Why Do Headphones Sound Good? : Sound Quality of Headphones

Sharon McEvilly
Aalto University
Master's Programme CCIS / AAT

sharon.mcevilly@aalto.fi

Abstract

Sound quality is an assessment of the accuracy, fidelity or intelligibility of audio output from an electronic device. Quality can be measured objectively, such as when tools are used to gauge the accuracy with which the device reproduces an original sound; or it can be measured subjectively, such as when human listeners respond to the sound or gauge its perceived similarity to another sound. In this report it will be taking a deeper dive into the investigation of the sound quality of headphones, how it is measured, factors that can sway a listener's perception of sound quality and how the measuring the frequency response of headphones and different to measuring the that of loudspeakers and if it is possible that they could be similar. This will be achieved by looking at various reports and papers written about the topic, going in-depth about the ways listening tests are done and relating this back to the basics of sound quality. Headphone sound quality is usually a measurement of accuracy and enjoyability. Performers in recording studios will often wear headphones to ensure that they can hear the other tracks they are singing with, without compromising the sound of their own recorded voice, so good sound quality in headphones is extremely important.

1 Introduction

As mentioned in the abstract of this paper, sound quality is an assessment of the accuracy, fidelity or intelligibility of audio output from an electronic device. However, when it comes to headphones specifically, the idea of sound quality is something that is widely debated. What constitutes "good" sound quality when we talk about headphones? Is it how much bass the headphones can emit, is it how the overall perception of the sound and one frequency level isn't overpowered compared to the

others? Or is it neither of these, and is it purely scientific based of what the frequency response is on the headphones and can only be determined through vigorous testing? This is what we will be investigating in this paper.

I have always been interested in headphones and finding a good pair of quality headphones as I am studying music technology in university, so music production is a big part of my work, so having a "good" pair of headphones is integral to my studies. Learning about how the sound quality of headphones is measured, learning about the different types of listening tests that are done when testing these headphones and learning about the frequency response of headphones and how that compares to other things, like speakers would be super useful to me so naturally this is a topic that has really caught my attention.

The paper will be discussing the types of listening tests, both objective and subjective, what they mean and the advantages and disadvantages of both. Perception is also a big factor when it comes to the sound quality of a set of headphones. Perception can be defined as "the ability to see, hear, or become aware of something through the senses" or "the way something is regarded, understood or interpreted." It is safe to say that perception is something that will play a big part in comparing the quality of headphones to each other and the different factors that could change perception will be discussed further on.

2 Sound Quality Evaluation for Headphones

Sound, as a whole, is a very complex matter, what we hear and measure is not the same. A humans perception can really sway the way we evaluated the quality of sound so how can we do this in an objective manner when evaluating the sound quality of headphones?

When evaluating the sound quality of headphones, one really needs to keep the perception a person might have in consideration, as it has been said: "Sound quality evaluation of headphones is mostly dependent on psychoacoustic listening tests." (H. Møller et al., 1995). While a highly subjective area, "The objective evaluation take less time and effort, and gives more repeatability." (H. Sung et al., 2007). In the paper written by H.Sung et al.,they set out to "introduce the difference between loudspeaker and headphone by analysing the physical data in view of sound quality" by carrying out the data analysis of headphones and then comparing it with that of a loudspeaker to find the differences and similarities to give them an idea on what a good headphone response should look like.

The tests carried out in this paper were done on a dummy head, possibly to get more reliable results as there would be many outside factors that could affect their results, which we will touch on later.H. Sung states "A full range MLS (Maximum Length Sequence) signal is generated and played by the headphone under test through the headphone amplifier (J. Browick, 2001). They collected the impulse response data

"By correlating the input and output MLS signal." It is also said that any test signal can be used as long as the impulse response can be obtained and with that, the impulse and frequency response are used to create the proposed evaluation system design.

2.1 Their findings

Sung et al. state in their paper that a good frequency has a "wide peak around 3kHz whereas perceptual responses of these headphones are supposed to be flat. This wide peak s caused by the ear canal effect while small fluctuations at higher frequencies over 6kHz are by the pinna where the personal dependency is dominant.

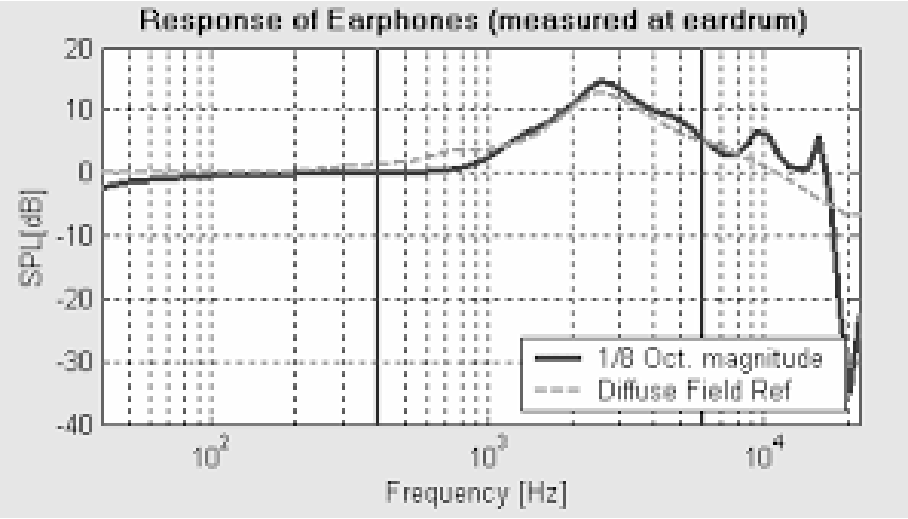

Figure 1: *Response of Headphones (measured at eardrum)*

In figure 1, it shows the typical response of headphones and Sung goes on to say that a loudspeaker "is regarded as ideal when it has a flat frequency response in a free field condition because it reproduces the same sound as recorded at the microphone", however, this differs to headphones as "the sound reproduced by headphone reaches the eardrum only across the ear canal which does not reflect natural listening environment". From their work in this paper, they can categorise the metrics needed for sound quality evaluation into 3 items: Tonal balance, Clarity and Bass Performance

2.2 Dummy head or human head?

Now that the 3 metrics that can be used for sound quality evaluation have been established, it feels like the next natural question would be: should we be testing on a dummy head or a real human and what are the benefits and disadvantages of both?

There can be arguments made for both sides. With an actual person being used to conduct the listening tests, the human listening system is very complex, therefore, we can get more complex and in-depth results which can lead to more concrete conclusions but you will always have the problem of bias and human perception. With a dummy head, the tests would be more effortless, less time consuming, and easier to repeat however, you are missing out on the complexity of the human hearing system so you might not get as in-depth results.

Objectively it is easier to use a dummy head for these tests. As mentioned by Günther Thiele, "The dummy head can be standardised. Results of different laboratories should have maximum conformity." and "The dummy head does not require insertion of a probe; there are no problems with the danger of hurting the ear drum and with hygiene." (Thiele, 1986). Thiele goes on to mention that "the existing dummy heads so not replicate important acoustical and mechanical properties of the human head, which are essential when coupling a headphone to the auditory canal." So how is this fixed? Thiele explains "Schröter and Eis have developed a special dummy head suitable for the measurement of earmuffs as well as earplugs" (J. Schröter and H. Eis, 1982). With this new dummy head, they can replicate the mechanical characteristics of the human head at the area of contact between the protector and the head.

So with this, the problem of missing out on the acoustical and mechanical properties of the human head so it would be more likely for a smoother process, researchers would use a dummy head, but a human head can still be used and offer pretty good results.

3 Factors that can sway perception

Now that some standards have been set for evaluating the sound quality of headphones, we move on to some factors that could possibly sway the results if the listening tests are conducted on humans.

When doing a listening tests for headphones, there are many things (both acoustical and non-acoustical) that can affect the perception of the sound quality. One of the main acoustical factors that could possibly sway perception of the sound quality of a set of headphones is the program that the listener is listening to.

In a experiment conducted by Sean E. Olive and others, they wanted to see, how much this would actually affect the listeners perceived results. It is important to pick the right type of program for these tests because as Olive et al. mentions in his paper "the acoustic characteristics and properties of the recordings themselves can bias and influence the results" (International Telecommunications Union, 2015).

3.1 What happened during the test

For this listening test, the programs that were picked "covered a variety of music genres and included tracks that vary in bandwidth, spectral density, dynamics and instrumentation" (Olive et al., 2017). When it came to headphones, they included six different models from varying brands all ranging in the price from 13 to 990 which was done to make sure they covered a wide range of sound quality and possibly create interactions between the programs and headphone ratings. The MUSHRA test method was used during this test and the relative levels use for the virtual headphones were matched according to ITU-R BS 1770-4 loudness model.

The results that were gathered from the test and that are shown in the table below indicate "that the main effect was Headphone. Program had no significant effect or interactions that influenced listener preference ratings".

Most of the people who took part in this test had a generally good agreement on with headphones sounded the best. There was one outlier in the tests however other than that, the results are pretty much the same, giving a result of r = .92

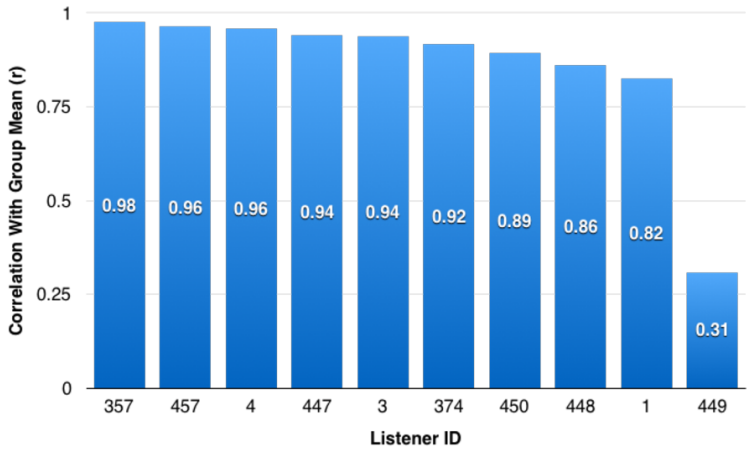

Figure 2: *Listener ID*

One could ask why did the program selected did not have an effect on the listeners perception as a lot of people have emotions tied to certain songs, Olive hypothesises in the paper "One explanation for this result is that the differences in headphone sound quality may have been sufficiently large to swamp out any program effects or interactions. Another explanation could be the programs were too similar in their physical characteristics to produce effects." So it is highly possible that in another test with a different set of headphones, the program could possibly have a big effect on the results.

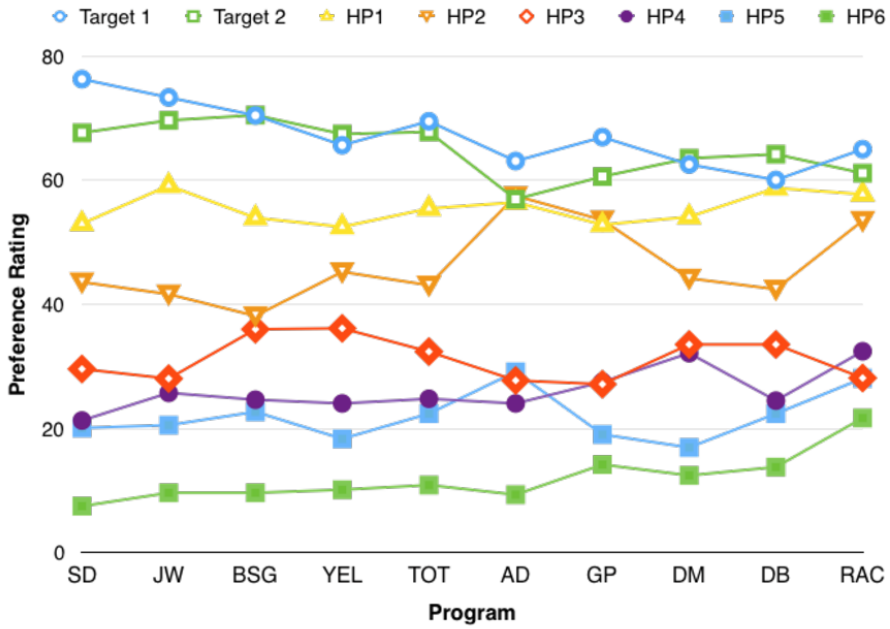

Figure 3: *Preferences in relation to program used*

3.2 Acoustical and Non-Acoustical biases

There are more factors that can influence a persons perception of the sound quality of headphones.

In a lecture given by Neo Kaplanis of Bang + Olufsen, he talks about some of the main biases that can be seen and even sometimes used in a listening test. Some of the Non-acoustical factors that Kaplanis talks about in this lecture are: 1)Visuals, 2)Smell and 3)Mood, Expectations and Familiarity.

To me, some of these factors make sense as with visuals, if a test subject likes the look of a headphone or sees that the headphones are branded, they will possibly have a preconceived perception that they are better if they were being compared to a not so well known brand or unbranded headphones. Mood also makes sense because a listener is happy when being in a test, they may like the headphones more and give them better results compared to if they were not in the best mood, they could possibly be more harsh in their reviews.

When it comes to the acoustical factors that can sway perceptions during listening test, there are a few that specific to headphones are wearables. Some of the factors that Kaplanis mentions are Head Placement, Head Size, Ear Size and probably most important, the overall fit of the headphones. If a pair of headphones do not fit a person, they are obviously going to be a bit more judgemental of the headphones than if they were to be a good overall fit and feel comfortable when wearing them.

Another factor that can also affect the overall fit is if the listener is wearing glasses or not so these things should be considered when choosing listener for these tests.

Perception can also be used to get a certain outcome in a listening test. An example of this that Kaplanis talked about during this lecture was when in 1915 Thomas Edison showed that his record player sounded identical to a live performer in 1915 and the results showed that the listeners agreed. In hindsight, it is known that this probably was not the case but how did he achieve this? Edison choose specific music that worked within the devices limitations so it would sound better, he trained the musicians to mimic the devices characteristics and he briefed the listeners before the test "Note that the voice of the artist and the voice of the Edison are indistinguishable." So with this example, you can really see that perception can have a lot of influence in how a test can turn out.

4 What makes a good pair of headphones?

It is a common perception that a pair of headphones from a brand or if they are expensive, they are "automatically" higher quality compared to cheaper and/or unbranded headphones. However, how true is that statement with what has been shown so far in this paper.

In a paper written by Sean E. Olive, he says "it seems that headphones designers are aiming at a target frequency response that is as random and variable as the weather." (Olive, 2022). He then goes on to talk about The International Electrotechnical Commission's (IEC) 60268-7 standard and how to measure the frequency response for free-field(FF) and diffuse-field(DF) and The International Telecommunication Union Radiocommunication Assembly (ITU-R) BS.708 standard recommends that professional headphones be designed to the DF target curve to achieve the best possible, however "most headphone designers have rejected this suggestion." Olive goes on to say that this is because "Recent pyschoacoustic investigators provide evidence that listeners prefer alternative headphone targets to DF and FF target standards" (Olive et al.2013b).

In this paper, they wanted to discover do listeners agree on what makes a headphone sound good and if this changes when it comes to age, gender or geographical location. In a similar study to this one (Oliver et al.,2013a) the listeners for this experiment said that "DF targets as having too much emphasis in upper midrange and lacking bass." They then went on to say that "The highest rated SRF target as having "good bass with an even spectral balance." So does this experiment hold out the same result??

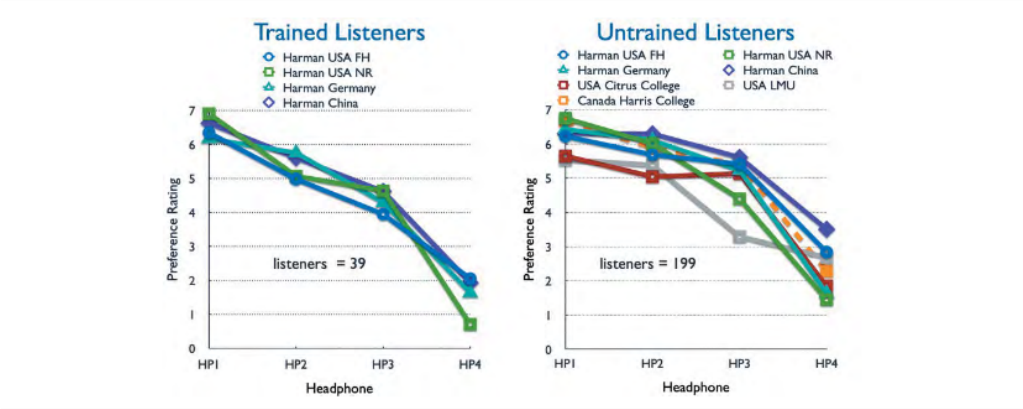

Figure 4: *Headphone Preferences*

4.1 The results

The preferences between the untrained and trained listeners were pretty consistent through all of the 11 tests as shown in the figure. They state that "the trained listeners were more discriminating and consistent than the untrained listeners.

In the study they have found out that female listeners preferred headphones with less bass compared to males, younger and less experienced listeners preferred more bass and treble compared to older and more experienced listeners and older (55+ years) listeners preferred more treble and less bass.

The results can also be showed through a table in relation to how listeners liked the sound of the headphones in relation to the Harman Target. The takeaway from this graph is that the Harman Target is a good indicator of what the general public, both trained and untrained listeners would like, but it is not advised to follow it strictly as it could alienate some people who enjoy more or less bass.

Distribution of Listener Categories Within Each Class (in %) Based on Preferred Headphone Sound Profile			
Category	**Class 1: Harman Target Lovers**	**Class 2: More Bass Is Better**	**Class 3: Less Bass Is Better**
Males	0.69	0.18	0.13
Females	0.56	0.04	0.40
Trained	0.70	0.30	0.00
Untrained	0.65	0.10	0.25
Age (years)			
20s	0.69	0.17	0.15
30s	0.74	0.13	0.13
40s	0.67	0.10	0.24
50+	0.30	0.20	0.50

Figure 5: *Headphone Preferences*

We can also see that in Figure 6 that they also looked to see if the retail price of a pair of headphones would correlate to the sound quality. From these graphs we can see that Around Ear headphones come the closest to the Harman Target and produce the highest ratings regardless of price. Over Ear headphones have the worst ratings and In Ear headphones are in the middle of the two. They also show that price is not a good indicator on if the sound quality is good or not and have pretty low correlation values.

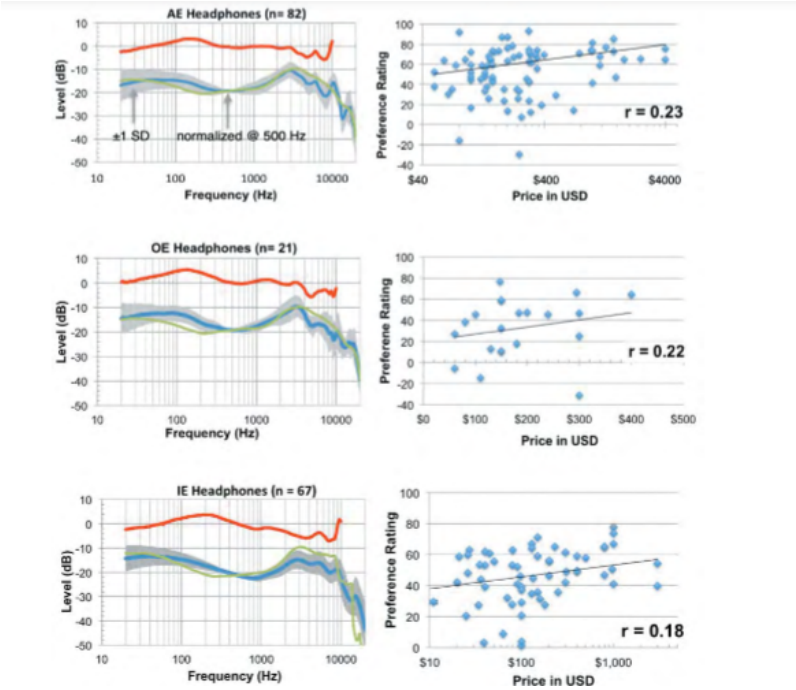

Figure 6: *Headphone Price versus Listener Preferences*

This sentiment is shared in a paper by Jeroen Breebaart where they also stated "no correlation could be observed between the measured magnitude response and retail price of headphones." (Breebaart, 2017). He did however find a interesting discovery in that "the variance in low frequency response seems to decrease with increasing price, indicating an improved bass response measurement consistency across headphones in the higher price range." So while, overall there is not a correlation between the price of headphones and sound quality, if you are looking for a better and consistent bass response, maybe it is worth investing.

5 Conclusion

In conclusion, there are a lot of things that need to be considered when looking into the sound quality of headphones. There are a lot of factors that need to be considered when you are planning to conduct a listening tests: are you going to use a dummy head or actual humans, what are the metrics you should use for testing

and what are some of the factors that can sway the perception of the listener in the test and how can use reduce these biases to get the fairest results. There is also no real correlation between how expensive a pair of headphones are and how good the sound quality and frequency response is, so that can be considered when you go to buy your next pair of headphones.

6 References

[1]H. Møller, C. B. Jensen, D. Hammershøi and M.F. sørensen , "Design Criteria for Headphones" J.Audio Eng. Soc. Vol. 43, No.4, pp.218-232(1995)

[2]Møller, H., Jensen, C.B., Hammershøi, D. and Sørensen, M.F., 1995. Design criteria for headphones. Journal of the Audio Engineering Society, 43(4), pp.218-232.

[3]J. Browick, Loudspeaker and headphone handbook 3rd Ed., Focal Press (2001)

[4]Theile, G., 1986. On the standardization of the frequency response of high-quality studio headphones. Journal of the Audio Engineering Society, 34(12), pp.956-969.

[5]

[6]Olive, S., Welti, T. and Khonsaripour, O., 2017, May. The influence of program material on sound quality ratings of in-ear headphones. In Audio Engineering Society Convention 142. Audio Engineering Society.

[7]Neo Kaplanis(2021). "Sound Quality Evaluation Theory and Practice" [Powerpoint Presentation], ELEC-E563001.

[8]Acoustics Today. 2022. The Perception and Measurement of Headphone Sound Quality- What Do Listeners Prefer? - Sean E. Olive. [online]

[9]S. Olive, T. Welti, and E. McMullin, "Listener Preferences for Different Headphone Target Response Curves," Paper 8867, (2013 May.).

[10]SE. E.. Olive, T. Welti, and E. McMullin, "A Virtual Headphone Listening Test Methodology," Paper 3-5, (2013 August.)

[11]Breebaart, J., 2017. No correlation between headphone frequency response and retail price. The Journal of the Acoustical Society of America, 141(6), pp.EL526-EL530.